4 Mathematical expectation

4a The definition

Discrete probability, in the elementary case of a *finite* probability space Ω, defines expectation of a random variable $X : \Omega \to \mathbb{R}$ in two equivalent ways:

$$\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega)p(\omega)\,, \tag{4a1}$$

$$\mathbb{E}(X) = \sum_{x} x\mathbb{P}\big(X = x \big)\,; \tag{4a2}$$

the latter sum is taken over the finite set of (possible) values of X. Of course, $\mathbb{P}\big(X = x \big) = \sum_{\omega:X(\omega)=x} p(\omega)$, and $p(\omega) = P(\{\omega\})$. Note basic properties of expectation:

4a3. If X and Y are identically distributed,[41] then $\mathbb{E}(X) = \mathbb{E}(Y)$.

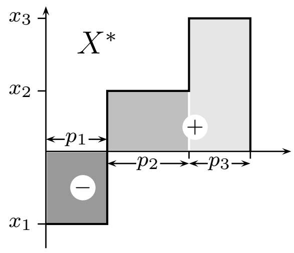
4a4. Monotonicity: if $X \le Y$ (that is, $\mathbb{P}\big(X \le Y \big) = 1$) then $\mathbb{E}(X) \le \mathbb{E}(Y)$.

4a5. Linearity:[42] $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ for all $a, b \in \mathbb{R}$.

Continuous probability does not use (4a1), and restricts (4a2) to the special case of 'simple' random variables X, that is, taking on a finite number of values. Properties 4a3–4a5 still hold for 'simple' random variables.[43]

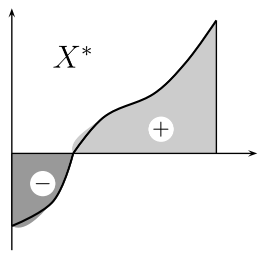
In terms of quantile functions, for 'simple' X,

$$\mathbb{E}(X) = \int_0^1 X^*(p)\,dp\,. \tag{4a6}$$



Expectation is just the average of all quantiles! It is natural to use the same integral for *defining* the expectation for all bounded random variables X:

$$\mathbb{E}(X) = \int_0^1 X^*(p)\,dp\,. \tag{4a7}$$



[41]These X, Y may be defined on different finite probability spaces.

[42]The set of all random variables on Ω is an n-dimensional linear space (here $n = |\Omega|$), and $\mathbb{E}(\cdot)$ is a linear functional on the space. Note that every X is a linear combination of indicators, and $\mathbb{E}(\mathbf{1}_A) = P(A)$; thus, $\mathbb{E}(\cdot)$ extends $P(\cdot)$ by linearity.

[43]These form a linear space, usually infinite-dimensional, and $\mathbb{E}(\cdot)$ is still a linear functional, extending $P(\cdot)$ by linearity.

Though, X^* may be bizarre, in particular, it need not be piecewise continuous (recall 2b6); but anyway, it is monotone and bounded, therefore Riemann integrable.[44] Properties 4a3–4a5 hold for all bounded random variables. Here are some hints toward proofs: 4a3: $F_X = F_Y$, therefore $X^* = Y^*$; 4a4: $F_X \ge F_Y$, therefore $X^* \le Y^*$; 4a5 ("sandwich argument"): given X and $\varepsilon > 0$, we may construct 'simple' $X_{\text{low}}, X_{\text{upp}}$ such that $X_{\text{low}} \le X \le X_{\text{upp}}$ and $X_{\text{upp}} - X_{\text{low}} \le \varepsilon$ everywhere; namely,

$$X_{\text{low}}(\omega) = n\varepsilon\,, \quad X_{\text{upp}}(\omega) = (n+1)\varepsilon \quad \text{whenever } X(\omega) \in [n\varepsilon, (n+1)\varepsilon)\,.$$

It follows that $|\mathbb{E}\,(X+Y) - \mathbb{E}\,(X) - \mathbb{E}\,(Y)| \le \varepsilon$ for all ε, therefore, $\mathbb{E}\,(X+Y) = \mathbb{E}\,(X) + \mathbb{E}\,(Y)$. It holds despite the fact that generally $(X+Y)^* \ne X^* + Y^*$ (try it for discrete X, Y).

No other extension of $\mathbb{E}\,(\cdot)$ from 'simple' to bounded random variables satisfies 4a4. In that sense, the definition (4a7) (or equivalent) is inescapable, as far as we do not want to lose 4a4 and (4a2).

If X is unbounded (I mean, $\mathbb{P}\big(a \le X \le b\big) \ne 1$ for all a, b) then X^* is also unbounded, therefore, not Riemann integrable, and we turn to improper integrals. Namely, if X is bounded from below but unbounded from above,[45] then X^* is bounded near 0 but unbounded near 1, and we define[46]

$$\mathbb{E}\,(X) = \lim_{\varepsilon\to 0} \int_0^{1-\varepsilon} X^*(p)\,dp \quad \in (-\infty, +\infty]\,. \tag{4a8}$$

The limit exists due to monotonicity, but may be infinite. If the limit is finite then X is called *integrable*, and $\mathbb{E}\,(X)$ is a number. If the limit is $+\infty$ then X is called non-integrable, and $\mathbb{E}\,(X) = +\infty$.

If X is bounded from above but unbounded from below, the situation is similar:

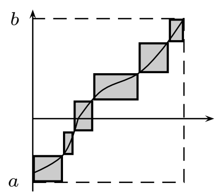
$$\mathbb{E}\,(X) = \lim_{\varepsilon\to 0} \int_\varepsilon^{1} X^*(p)\,dp \quad \in [-\infty, +\infty)\,. \tag{4a9}$$

If X is unbounded both from above and from below, the situation is more complicated. Expectation is *not* defined as $\lim_{\varepsilon\to 0} \int_\varepsilon^{1-\varepsilon} X^*(p)\,dp$; rather, it is defined by

$$\mathbb{E}\,(X) = \lim_{\varepsilon\to 0} \int_\varepsilon^{p_0} X^*(p)\,dp + \lim_{\varepsilon\to 0} \int_{p_0}^{1-\varepsilon} X^*(p)\,dp\,, \tag{4a10}$$

where $p_0 \in (0,1)$ may be chosen arbitrarily (it does not affect the result). If both limits are finite, then X is called integrable, and $\mathbb{E}\,(X)$ is a number. Otherwise X is called non-integrable. If the first limit is finite while the second is infinite, then $\mathbb{E}\,(X) = +\infty$. If the

[44] Hint toward the proof: the difference of two 'simple' integrals (upper and lower) does not exceed $(b-a)/n$ (irrespective of (dis)continuity of X^*).



[45] Sets of zero probability may be neglected, as usual.

[46] An equivalent definition: $\mathbb{E}\,(X) = \sup\{\mathbb{E}\,(Y) : Y \text{ is 'simple'}, Y \le X\}$.

first limit is $(-\infty)$ while the second limit is finite then $\mathbb{E}(X) = -\infty$. If the first limit is $(-\infty)$ while the second is $(+\infty)$ then $\mathbb{E}(X)$ is not defined (never write $\mathbb{E}(X) = \infty$ in that case).

The formula $\mathbb{E}(X) = \lim_{\varepsilon\to 0} \int_\varepsilon^{1-\varepsilon} X^*(p)\,dp$ gives correct answers in three cases (number + number = number; number + $(+\infty) = +\infty$; $(-\infty)$ + number $= -\infty$) but can give a wrong answer in the fourth case ($(-\infty) + (+\infty)$ is undefined).

4a11 Definition. (a) A random variable X is *integrable,* if its quantile function X^* is an integrable function on $(0,1)$.

(b) *Expectation* of an integrable random variable is the integral of its quantile function,

$$\mathbb{E}(X) = \int_0^1 X^*(p)\,dp\,.$$

Riemann integration is meant in (a) and (b), improper integral being used when needed, see (4a8)–(4a10).

Arbitrariness of the quantile function (at jumps) does not matter. Additional cases $\mathbb{E}(X) = -\infty$, $\mathbb{E}(X) = +\infty$, not stipulated by the definition, are also used. Note that X is integrable if and only if $\mathbb{E}|X| < \infty$.[47] Properties 4a3–4a5 still hold. Namely:

4a12. If X and Y are identically distributed, then $\mathbb{E}(X) = \mathbb{E}(Y)$ provided that both X and Y are integrable; otherwise both X and Y are not integrable.

4a13. Monotonicity: if X, Y are integrable and $X \le Y$ (that is, $\mathbb{P}\big(X \le Y\big) = 1$) then $\mathbb{E}(X) \le \mathbb{E}(Y)$.

4a14. Linearity: if X, Y are integrable then $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ for all $a, b \in \mathbb{R}$.

Here is a useful inequality: for every event A,

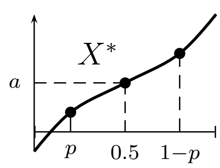
(4a15) $$\int_0^{\mathbb{P}(A)} X^*(p)\,dp \le \mathbb{E}(\mathbf{1}_A X) \le \int_{1-\mathbb{P}(A)}^1 X^*(p)\,dp\,.$$

4b Expectation and symmetry

4b1 Definition. A random variable X (and its distribution) is *symmetric* with respect to a number $a \in \mathbb{R}$ (called the symmetry center) if random variables X and $2a - X$ are identically distributed.

In terms of the quantile function,

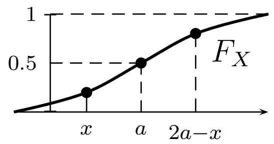
$$X^*(p) + X^*(1-p) = 2a$$



[47] We cannot say that X is integrable if and only if $|\mathbb{E}\,X| < \infty$ (think, why).

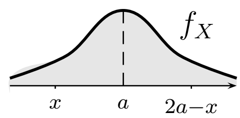
(except for jumping points, maybe; in any case, $X^*(p+) + X^*((1-p)-) = 2a$). By the way, it follows that X cannot have more than one symmetry center. In terms of the distribution function,

$$F_X(x) + F_X\big((2a-x)-\big) = 1\,.$$



In terms of the density (if it exists),

$$f_X(x) = f_X(2a-x)$$



for all $x \in \mathbb{R}$ (except for a set of measure 0, maybe).

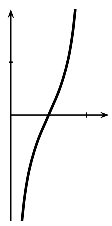
4b2 Exercise. Show that the bizarre random variable Y of Example 2b6 is symmetric; what is its symmetry center? Hint: $(0.101010\dots)_2 - (0.\beta_1 0\beta_3 0\beta_5 0\dots)_2 = (0.\overline{\beta}_1 0\overline{\beta}_3 0\overline{\beta}_5 0\dots)_2$.

4b3 Exercise. If X is symmetric w.r.t. a and integrable then

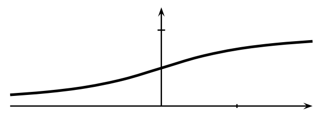
$$\mathbb{E}\,(X) = a\,.$$

Prove it. Hint: $\mathbb{E}\,(X) = \mathbb{E}\,(2a - X)$.

You may ask: why bother about integrability? If X is symmetric w.r.t. a, what is really wrong in saying that $\mathbb{E}\,(X) = 0$ by symmetry, without checking integrability? Well, let us perform a numeric experiment. Consider so-called Cauchy distribution:



$$X^*(p) = \tan\pi\big(p - \tfrac12\big)\,; \qquad F_X(x) = \tfrac12 + \tfrac1\pi \arctan x\,.$$



It is symmetric w.r.t. 0. The theory says that X is non-integrable, while $Y = \sqrt[3]{X}$ is integrable (and still symmetric w.r.t. 0). Take a sample of, say, 1000 independent values of X. To this end, we take 1000 independent random numbers $\omega_1, \dots, \omega_{1000}$ according to the uniform distribution on $(0,1)$, and calculate $x_k = \tan\pi(\omega_k - \frac12)$; here is my sample (sorted):

k	1	2	3	4	...	997	998	999	1000
ω_k	.000703	.001391	.002178	.002693	...	.991265	.992999	.995051	.998198
x_k	-452.5	-228.7	-146.1	-118.1	...	36.4	45.4	64.3	176.6
$\sqrt[3]{x_k}$	-7.677	-6.115	-5.267	-4.907	...	3.315	3.568	4.006	5.610

Calculate mean values:

$$\frac{x_1 + \dots + x_{1000}}{1000} = -1.23; \qquad \frac{\sqrt[3]{x_1} + \dots + \sqrt[3]{x_{1000}}}{1000} = 0.0021$$

You see, the mean of $\sqrt[3]{x_k}$ is close to 0, which cannot be said about the mean of x_k. The experiment confirms the theory; $\mathbb{E}\,\sqrt[3]{X} = 0$, but $\mathbb{E}\,X$ is undefined. In fact, it is well-known that $\frac1n(X_1 + \dots + X_n)$ has the Cauchy distribution, again (with no scaling).

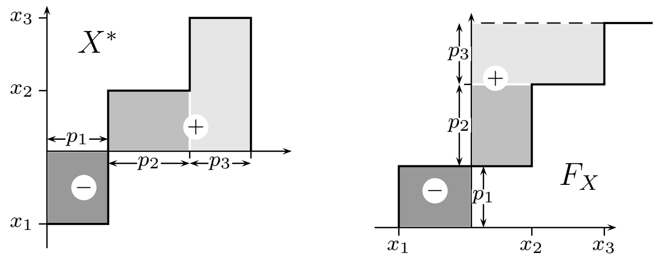
4c Equivalent definitions

For ‘simple’ X, (4a6) is easy to rewrite in terms of the distribution function:

$$\mathbb{E}\,(X) = \int_0^1 X^*(p)\,dp =$$

$$(4c1)\qquad = -\int_{-\infty}^0 F_X(x)\,dx + \int_0^\infty \bigl(1 - F_X(x)\bigr)\,dx =$$

$$(4c2)\qquad = -\int_{-\infty}^0 \mathbb{P}\,\bigl(\,X < x\,\bigr)\,dx + \int_0^\infty \mathbb{P}\,\bigl(\,X > x\,\bigr)\,dx\,.$$



(No matter, whether we use $\mathbb{P}\,\bigl(\,X < x\,\bigr)$ or $\mathbb{P}\,\bigl(\,X \le x\,\bigr)$.)

As before, a limiting procedure generalizes formulas (4c1), (4c2) to all bounded random variables, and another limiting procedure generalizes them to unbounded random variables. The latter case needs improper Riemann integrals. Still, we have 4 cases: bounded; bounded from below only; bounded from above only; unbounded both from above and from below. The latter case leads to

$$(4c3)\qquad \mathbb{E}\,(X) = -\lim_{M\to\infty}\int_{-M}^0 F_X(x)\,dx + \lim_{M\to\infty}\int_0^M \bigl(1 - F_X(x)\bigr)\,dx\,;$$

if both terms are finite then X is integrable, otherwise it is not. Additional cases $\mathbb{E}\,(X) = -\infty$, $\mathbb{E}\,(X) = +\infty$ are treated as before.

One more equivalent definition will be given, see 4d10.

Assume that X has a density f_X. Then (see (4a2), see also 4d15 in the next subsection)

$$(4c4)\qquad \mathbb{E}\,(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx\,.$$

Riemann integration is enough if X is bounded *and* f_X is Riemann integrable. Improper Riemann integral is used for unbounded X and/or f_X with singularities (recall 2c3). In principle, f_X may be too bizarre for Riemann integration (see page 16), then Lebesgue integration must be used in (4c4).

Integrability of X is equivalent to convergence of both limits in the next formula:

$$(4c5)\qquad \mathbb{E}\,(X) = \lim_{M\to\infty}\int_{-M}^0 x f_X(x)\,dx + \lim_{M\to\infty}\int_0^M x f_X(x)\,dx\,.$$

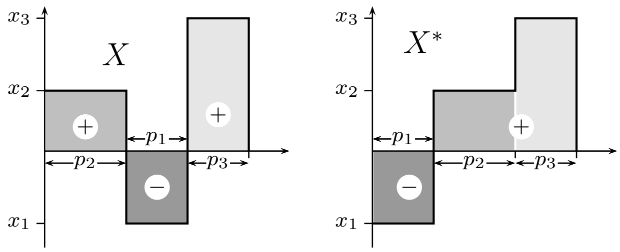
4d Expectation and Lebesgue integral

Probably you think that Lebesgue integral is a very complicated notion. Here comes a surprise: in some sense, you are *already* acquainted with Lebesgue integration, since it is basically the same as expectation![48]

Let our probability space $(\Omega, \mathcal{F}, P)$ be $(0,1)$ with Lebesgue measure on Borel σ-field. Then a random variable X is just a Borel function $X : (0,1) \to \mathbb{R}$ (recall 3d1).

Let X be a step function, then clearly

(4d1)
$$\mathbb{E}\,(X) = \int_0^1 X^*(p)\,dp = \int_0^1 X(p)\,dp\,.$$



You see, X^* is an *increasing rearrangement* of X, and so, their mean values are equal. A sandwich argument shows that (4d1) holds for all functions X Riemann integrable on $[0,1]$. Does it hold for bizarre Borel functions X? Of course, it does not; X^*, being monotone, is within the reach of Riemann integration, but X is generally beyond it.[49] However, we may *define* $\int_0^1 X(p)\,dp$ as $\int_0^1 X^*(p)\,dp$, and this is just Lebesgue integration!

4d2 Definition. (a) A Borel function $\varphi : (0,1) \to \mathbb{R}$ is (Lebesgue) *integrable,* if

$$\int_{-\infty}^0 \operatorname{mes}\{x \in (0,1) : \varphi(x) < y\}\,dy < \infty\,,$$
$$\int_0^{+\infty} \operatorname{mes}\{x \in (0,1) : \varphi(x) > y\}\,dy < \infty\,.$$

(b) (Lebesgue) *integral* of an integrable function $\varphi : (0,1) \to \mathbb{R}$ is

$$\int_0^1 \varphi(x)\,dx = -\int_{-\infty}^0 \operatorname{mes}\{x \in (0,1) : \varphi(x) < y\}\,dy + \int_0^{+\infty} \operatorname{mes}\{x \in (0,1) : \varphi(x) > y\}\,dy\,.$$

So, (4d1) holds[50] for all Borel functions $X : (0,1) \to \mathbb{R}$ provided that $\int_0^1 X(p)\,dp$ is treated as Lebesgue integral. Lebesgue integral extends proper and improper Riemann integrals. Thus, $\int_0^1 X^*(p)\,dp$ may be treated equally well as Riemann or Lebesgue integral. Properties of expectation become properties of Lebesgue integral:

[48]However, this section is not a substitute of *measure theory* course. Especially, we borrow from measure theory existence of Lebesgue measure, some properties of Lebesgue integral, etc.

[49]An example: the indicator of the set of rational numbers.

[50]That is, the four cases (a number, $-\infty$, $+\infty$, $\infty - \infty$) for $\int_0^1 X^*(p)\,dp$ correspond to the four cases for $\int_0^1 X(p)\,dp$, and in the first case the two integrals are equal.

4d3. Monotonicity: if $X(\omega) \le Y(\omega)$ for all[51] $\omega \in (0,1)$ then $\int_0^1 X(\omega)\,d\omega \le \int_0^1 Y(\omega)\,d\omega$.

4d4. Linearity: $\int_0^1 \big(aX(\omega) + bY(\omega)\big) = a\int_0^1 X(\omega)\,d\omega + b\int_0^1 Y(\omega)\,d\omega$.

Definition 4d2 generalizes readily for a Borel function φ on a Borel set $B \subset \mathbb{R}^d$, $d = 1,2,3,\dots$ (see 5a14); namely,

$$\int_B \varphi(x)\,dx = -\int_{-\infty}^0 \operatorname{mes}_d\{x \in B : \varphi(x) < y\}\,dy + \int_0^{+\infty} \operatorname{mes}_d\{x \in B : \varphi(x) > y\}\,dy\,; \tag{4d5}$$

as before, mes_d stands for the d-dimensional Lebesgue measure (see 1f12); $\operatorname{mes}_d(B)$ may be any number or $+\infty$.[52] Monotonicity and linearity still hold.

4d6 Exercise. Prove that

$$\int_B \mathbf{1}_A(x)\,dx = \operatorname{mes}_d(A)\,;$$

here $\mathbf{1}_A$ is the indicator of a Borel subset A of the Borel set $B \subset \mathbb{R}^d$. What happens if $\operatorname{mes}_d(A) = 0$?

Lebesgue integral over an arbitrary probability space[53] $(\Omega, \mathcal{F}.P)$ is defined similarly,

$$\int_\Omega X\,dP = -\int_{-\infty}^0 P\{\omega \in \Omega : X(\omega) < x\}\,dx + \int_0^{+\infty} P\{\omega \in \Omega : X(\omega) > x\}\,dx\,; \tag{4d7}$$

it is also written as $\int_\Omega X(\omega)\,dP(\omega)$ or $\int_\Omega X(\omega)\,P(d\omega)$. So,

$$\mathbb{E}\,(X) = \int_\Omega X\,dP\,, \tag{4d8}$$

which is a continuous counterpart of (4a1). Now we may rewrite $\mathbb{E}\,(X+Y) = \mathbb{E}\,(X) + \mathbb{E}\,(Y)$ in the form $\int_\Omega (X+Y)\,dP = \int_\Omega X\,dP + \int_\Omega Y\,dP$ (though, it is just another notation for the same statement).

In particular, let P be a one-dimensional distribution (that is, a probability measure on $(\mathbb{R}, \mathcal{B})$, see Sect. 2d), and $F = F_P$ its cumulative distribution function (that is, $F(x) = P\big((-\infty, x]\big)$. Then we write

$$\int_{\mathbb{R}} X\,dP = \int_{\mathbb{R}} X\,dF = \int_{-\infty}^{+\infty} X(\omega)\,dF(\omega) \tag{4d9}$$

for any Borel function $X : \mathbb{R} \to \mathbb{R}$. Analysts prefer to write something like $\int_{-\infty}^{+\infty} \varphi(x)\,dF(x)$, but it is the same. Such integrals are called *Lebesgue-Stieltjes integrals.* Note that F need not be differentiable, moreover, it need not be continuous.

Here is another equivalent definition of expectation:

$$\mathbb{E}\,(X) = \int_{-\infty}^{+\infty} x\,dF_X(x)\,. \tag{4d10}$$

---

[51] Or 'almost all', that is, except of a set of measure 0.

[52] It may happen that $\operatorname{mes}_d\{x \in B : \varphi(x) < y\} = +\infty$ for some $y > 0$; then, of course, $\int_0^{+\infty} \operatorname{mes}_d\{x \in B : \varphi(x) > y\}\,dy = +\infty$. The same for the negative part.

[53] More generally, an arbitrary *measure space.*

4d11 Exercise. Prove (4d10). Hint: define a random variable X_* on the probability space $(\mathbb{R}, \mathcal{B}, P_X)$ by $X_*(\omega) = \omega$ (see page 19), then X_* and X are identically distributed.

4d12 Exercise. Simplify $\int_{-\infty}^{+\infty} \varphi(x)\, dF(x)$ assuming that F is a step function (while φ is just a Borel function). Can a single value of φ influence the integral?

4d13 Exercise. Show that (4d10) boils down to (4a2) when F_X is a step function.

4d14 Proposition. If P has a density f then

$$\int_{\mathbb{R}} \varphi\, dP = \int_{-\infty}^{+\infty} \varphi(x) f(x)\, dx$$

for any Borel function $\varphi : \mathbb{R} \to \mathbb{R}$. Both integrals are Lebesgue integrals. The four cases (a number, $-\infty$, $+\infty$, $\infty - \infty$) for the former integral correspond to the four cases for the latter integral.

4d15 Exercise. Derive (4c4) from (4d10) and 4d14.

If F is smooth, we have $f = F'$, and (4d14) becomes

$$\int_{-\infty}^{+\infty} \varphi(x)\, dF(x) = \int_{-\infty}^{+\infty} \varphi(x)\, F'(x)\, dx\,.$$

4e Expectation and transformations

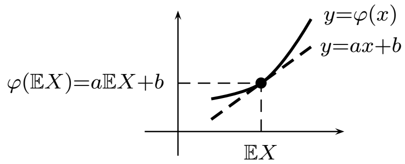
Let X be an integrable random variable and $Y = \varphi(X)$ its transformation (see Sect. 3). If φ is a linear function, then Y is also integrable and

(4e1) $$\mathbb{E}\, Y = \varphi(\mathbb{E}\, X) \qquad \text{(for linear } \varphi)$$

Indeed, $\varphi(x) = ax + b$; $Y = aX + b$; $\mathbb{E}\, Y = a\mathbb{E}\, X + b = \varphi(\mathbb{E}\, X)$. If φ is a *convex* function, then[54]

(4e2) $$\mathbb{E}\, Y \ge \varphi(\mathbb{E}\, X)\,, \qquad \text{(for convex } \varphi)$$

which is well-known as *Jensen inequality.* Hint to a proof:



$$\mathbb{E}\, Y = \mathbb{E}\, \varphi(X) \ge \mathbb{E}\,(aX + b) = a\mathbb{E}\, X + b = \varphi(\mathbb{E}\, X)\,.$$

Similarly,

(4e3) $$\mathbb{E}\, Y \le \varphi(\mathbb{E}\, X) \qquad \text{(for concave } \varphi).$$

[54]Though, it can happen that Y is non-integrable; in that case $\mathbb{E}\, Y = +\infty$.

In principle, for any Borel function φ, the expectation $\mathbb{E}\,Y = \mathbb{E}\,\varphi(X)$ may be calculated by applying any definition (4a11, (4c1), (4d10), maybe (4c4)) to Y. However, calculating Y^*, F_Y or f_Y can be tedious, if φ is non-monotone (see Sect. 3c). Fortunately, $\mathbb{E}\,\varphi(X)$ can be calculated via X^*, F_X or f_X; namely,

$$\mathbb{E}\,\varphi(X) = \int_0^1 \varphi\big(X^*(p)\big)\,dp\,; \tag{4e4}$$

$$\mathbb{E}\,\varphi(X) = \int_{-\infty}^{+\infty} \varphi(x)\,dF_X(x)\,; \tag{4e5}$$

$$\mathbb{E}\,\varphi(X) = \int_{-\infty}^{+\infty} \varphi(x) f_X(x)\,dx \qquad \text{(if } f_X \text{ exists).} \tag{4e6}$$

Four possible cases (a number, $-\infty$, $+\infty$, $\infty-\infty$) appear simultaneously in these expressions.

Formula (4e4) follows from 2e9 and 3d14; X and X^* are identically distributed, therefore $\varphi(X)$ and $\varphi(X^*)$ are identically distributed, therefore their expectations are equal.

Formula (4e5) is obtained similarly, but X^* is replaced with the random variable X_* defined on the probability space $(\mathbb{R}, \mathcal{B}, P_X)$ by $X_*(\omega) = \omega$.

Formula (4e6) follows from (4e5) and 4d14.

If X is bounded, $\mathbb{P}\big(a \le X \le b\big) = 1$, then of course (4e4)–(4e6) hold for any Borel function $\varphi : [a,b] \to \mathbb{R}$ (rather than $\mathbb{R} \to \mathbb{R}$). Similarly, if we have a Borel set $B \subset \mathbb{R}$ such that $\mathbb{P}\big(X \in B\big) = 1$, then it is enough for φ to be defined on B. Say, $\mathbb{E}\sqrt{X}$ may be treated, provided that $\mathbb{P}\big(X \ge 0\big) = 1$.

Note that probabilities are never transformed, only values; never write something like $\varphi(p)$, or $\varphi\big(F_X(x)\big)$, or $\varphi\big(f_X(x)\big)$.

Especially, functions $|x| = x^+ + x^-$, $x^+ = \max(0,x)$ and $x^- = \max(0,-x)$ lead to

$$\begin{aligned}
&\mathbb{E}\,|X| = \int_0^1 |X^*(p)|\,dp = \int_{-\infty}^{+\infty} |x|\,dF_X(x) = \int_{-\infty}^{+\infty} |x| f_X(x)\,dx\,,\\
&\mathbb{E}\,X^+ = \int_0^1 \big(X^*(p)\big)^+\,dp = \int_0^{\infty} x\,dF_X(x) = \int_0^{\infty} x f_X(x)\,dx\,,\\
&\mathbb{E}\,X^- = \int_0^1 \big(X^*(p)\big)^-\,dp = \int_{-\infty}^{0} (-x)\,dF_X(x) = \int_{-\infty}^{0} (-x) f_X(x)\,dx\,;\\
&|X| = X^+ + X^-\,; \qquad \mathbb{E}\,|X| = \mathbb{E}\,(X^+) + \mathbb{E}\,(X^-)\,;\\
&\mathbb{E}\,|X| < +\infty \quad\Longleftrightarrow\quad X \text{ is integrable}\,,\\
&\mathbb{E}\,|X| = +\infty \quad\Longleftrightarrow\quad X \text{ is non-integrable}\,.
\end{aligned}$$

4f Variance, moments, generating function

Variance and standard deviation are defined in continuous probability in the same way as in discrete probability:

$$\begin{aligned}
&\mathrm{Var}(X) = \mathbb{E}\,(X^2) - (\mathbb{E}\,X)^2 = \mathbb{E}\,\big(X - \mathbb{E}\,X\big)^2 = \min_{a\in\mathbb{R}} \mathbb{E}\,(X-a)^2\,;\\
&\sigma(X) = \sqrt{\mathrm{Var}(X)}\,.
\end{aligned} \tag{4f1}$$

Clearly,

$$(4f2) \qquad Y = aX + b \quad \Longrightarrow \quad \begin{array}{l} \operatorname{Var}(Y) = a^2 \operatorname{Var}(X)\,, \\ \sigma(Y) = |a|\sigma(X)\,. \end{array}$$

It may happen that X^2 is non-integrable; then $\mathbb{E}\,(X-a)^2 = +\infty$ for every a, and we write $\operatorname{Var}(X) = \infty$.

If X is non-integrable then X^2 is also non-integrable, which follows from such inequalities as $|x| \le \max(1, x^2)$ or $|x| \le \frac{1}{2}(1+x^2)$.

Moments:

$$(4f3) \qquad \begin{array}{ll} \mathbb{E}\, X^{2n} \in [0, +\infty]\,; & \\ \mathbb{E}\, X^{2n+1} \in \mathbb{R} & \text{if } \mathbb{E}\,|X|^{2n+1} < \infty\,. \end{array}$$

Moment generating function:

$$(4f4) \qquad \operatorname{MGF}_X(t) = \mathbb{E}\left(e^{tX}\right) \in (0, +\infty]\,.$$

You may easily express these in terms of X^*, or F_X, or f_X (if exists); the latter is especially useful:

$$(4f5) \qquad \begin{aligned} \mathbb{E}\, X^{2n} &= \int_{-\infty}^{+\infty} x^{2n} f_X(x)\, dx \in [0, +\infty]\,; \\ \mathbb{E}\, X^{2n+1} &= \int_{-\infty}^{+\infty} x^{2n+1} f_X(x)\, dx \in \mathbb{R} \qquad \text{if } \int_{-\infty}^{+\infty} |x|^{2n+1} f_X(x)\, dx < \infty\,; \\ \operatorname{MGF}_X(t) &= \int_{-\infty}^{+\infty} e^{tx} f_X(x)\, dx \in (0, +\infty]\,. \end{aligned}$$

Compare it with elementary formulas of Introduction to Probability: $\mathbb{E}\,(X^n) = \sum_k x_k^n p_k$, $\mathbb{E}\left(e^{tX}\right) = \sum_k e^{tx_k} p_k$.

4f6 Exercise. If $0 \le s \le t$ and $\operatorname{MGF}_X(t) < \infty$ then $\operatorname{MGF}_X(s) < \infty$. Prove it. Hint: $e^{sx} \le \max(1, e^{tx})$. What can you say about the case $t \le s \le 0$? Show that for any X, the set $\{t : \operatorname{MGF}_X(t) < \infty\}$ is an interval, containing 0. Can it be of the form $(a,b), [a,b], [a,b), (a,b]$? Can a, b be 0 or $\pm\infty$?

4f7 Proposition. If $\operatorname{MGF}_X(t) < \infty$ for all $t \in (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$, then $\mathbb{E}\,|X|^n < \infty$ for all n, and

$$\operatorname{MGF}_X(t) = 1 + t\mathbb{E}\,(X) + \frac{1}{2!} t^2 \mathbb{E}\,(X^2) + \frac{1}{3!} t^3 \mathbb{E}\,(X^3) + \dots$$

for all $t \in (-\varepsilon, \varepsilon)$.

For bounded X the proof is rather easy: $e^{tX} = 1 + tX + \frac{1}{2!} t^2 X^2 + \dots$ and the series converges uniformly. For unbounded X the convergence is non-uniform, but still, the expectations converge (by Dominated Convergence Theorem 6c13).

4f8 Example. The exponential distribution: $f_X(x) = e^{-x}$ for $x > 0$;

$$\mathbb{E}\, e^{tX} = \int_0^\infty e^{tx} e^{-x} \, dx = \int_0^\infty e^{-(1-t)x} \, dx \, ;$$

for $t < 1$ the integral converges to $\frac{1}{1-t}$,

$$\mathrm{MGF}_X(t) = \frac{1}{1-t} = 1 + t + t^2 + \dots$$

Comparing it with 4f7 we get

$$\mathbb{E}\, X^n = n!$$

4f9 Example. Let X be uniform on $(0,1)$, then $f_X(x) = 1$ for $x \in (0,1)$; $\mathbb{E}\, e^{tX} = \int_0^1 e^{tx} \, dx$;

$$\mathrm{MGF}_X(t) = \frac{e^t - 1}{t} = \frac{1}{t}\Big(-1 + 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots \Big) = 1 + \frac{t}{2!} + \frac{t^2}{3!} + \dots \, .$$

Comparing it with (4f7) we get

$$\mathbb{E}\, X^n = \frac{1}{n+1}$$

which is easy to find without MGF.

4f10 Example. The normal distribution:[55]

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, .$$

We have

$$\mathrm{MGF}_X(t) = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx = e^{t^2/2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} \, dx =$$

$$= e^{t^2/2} = 1 + \frac{t^2}{2} + \frac{1}{2!}\Big(\frac{t^2}{2}\Big)^2 + \dots$$

for all $t \in (-\infty, +\infty)$. Therefore X has all moments, and

$$\mathbb{E}\,(X^{2n}) = \frac{(2n)!}{2^n n!} = \frac{1 \cdot 2 \cdot 3 \cdot \dots \cdot (2n)}{2 \cdot 4 \cdot 6 \cdot \dots \cdot (2n)} = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n-1)$$

for $n = 1, 2, 3, \dots$; in particular,

$$\mathbb{E}\,(X^2) = 1 \, ; \quad \mathbb{E}\,(X^4) = 3 \, ; \quad \mathbb{E}\,(X^6) = 15 \, .$$

Of course, $\mathbb{E}\,(X^{2n-1}) = 0$ by symmetry.

4f11 Exercise. Let $\mathbb{P}\big(X > 0 \big) = 1$, and X have a density f_X such that $f_X(x) \sim \frac{\mathrm{const}}{x^\alpha}$ for $x \to +\infty$ (that is, $x^\alpha f_X(x)$ converges to a positive number when $x \to +\infty$; here α is a positive parameter). Prove that $\mathbb{E}\,(X^n) < \infty$ if and only if $n < \alpha - 1$. Apply it to Cauchy distribution (page 36).

[55] For now I do not explain, why $\int_{-\infty}^{+\infty} e^{-x^2/2} \, dx = \sqrt{2\pi}$. Strangely enough, that 'one-dimensional' fact is a natural by-product of two-dimensional theory.